First, I would like you to go ahead and navigate to **Section E**, or the data preview area of the Data Source Page. Remember from the last chapter that this is the area where you can review some of the rows of the dataset you are loading into the Tableau Workbook.

Project review_id	raw!data+ Country	raw!data+ Points	raw!data+ Price	raw!data+ Province	raw!data+ Region 1	raw!data+ Title	raw!data+ Variety	raw!data+ Winery	wine!experts Taster Name
1,558	France	98	1900,0	Bordeaux	Margaux	Château Margaux 20...	Bordeaux-style Red B...	Château Margaux	Roger Voss
1,566	France	97	1100,0	Bordeaux	Pessac-Léognan	Château La Mission H...	Bordeaux-style Red B...	Château La Mission H...	Roger Voss
1,571	France	96	1200,0	Bordeaux	Pessac-Léognan	Château Haut-Brion ...	Bordeaux-style Red B...	Château Haut-Brion	Roger Voss
1,575	France	96	1300,0	Bordeaux	Pauillac	Château Mouton Rot...	Bordeaux-style Red B...	Château Mouton Rot...	Roger Voss
15,840	France	96	2500,0	Bordeaux	Pomerol	Château Pétrus 2014...	Bordeaux-style Red B...	Château Pétrus	Roger Voss
30,110	France	97	886,0	Burgundy	Montrachet	Olivier Leflaive 2014 ...	Chardonnay	Olivier Leflaive	Roger Voss
36,529	France	99	800,0	Champagne	Champagne	Krug 2002 Clos du M...	Chardonnay	Krug	Roger Voss
36,531	Portugal	97	1000,0	Port	null	W. & J. Graham's NV ...	Port	W. & J. Graham's	Roger Voss
39,288	France	98	800,0	Bordeaux	Pessac-Léognan	Château Haut-Brion ...	Bordeaux-style Whit...	Château Haut-Brion	Roger Voss
41,441	Portugal	94	980,0	Port	null	Kopke 1935 Colheita ...	Port	Kopke	Roger Voss

Section E, the data preview area.
Remember, the default is to show the first 1000 rows of the data.

In our preview, we are reviewing the first 1000 rows of the resulting table from combining three different Excel sheets (in our case, it would be the three sheets in this Excel file). Tableau also defaults to the **data source** order. This means that it will sort rows based on the Excel sheet they were in. There are other options to sort the rows of data that you see in the data preview area. Click on the drop-down arrow for *Sort fields* (in the top-left area) to see the other options.
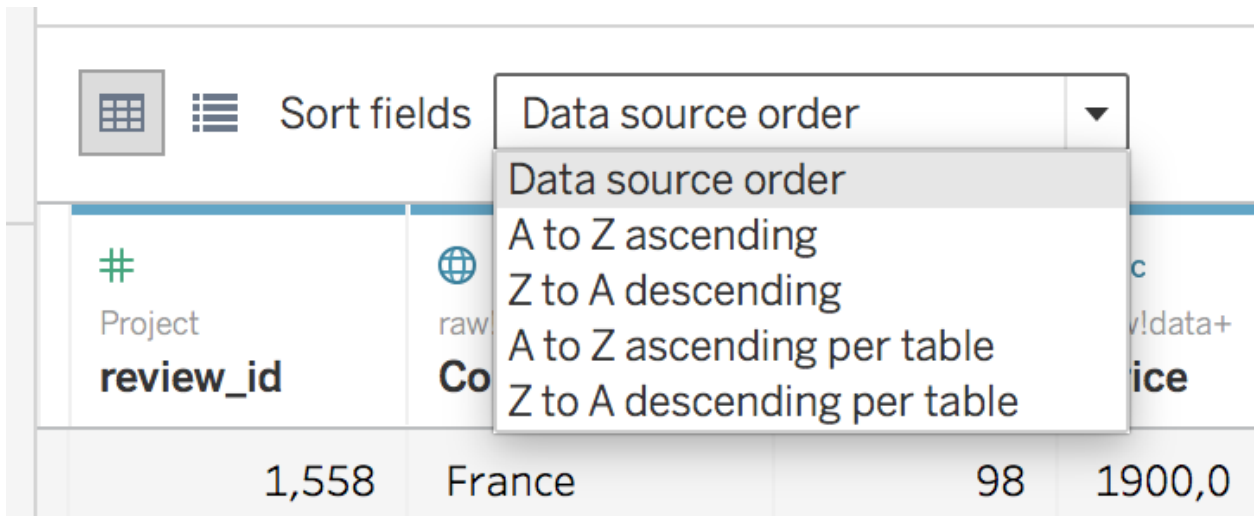
The drop-down menu for the Sort fields option in the top-left area of Section E.
Tableau does *not* change the original data files that are used to load data into the Tableau workbook. In other words, you *cannot* delete or create any columns or rows of data in the source file.

You can hide, or filter out, columns as well as create new columns and calculated fields in the Tableau workbook, but those changes are *not* reflected in the associated **data source** file (in our case, the Excel file linked above).

PRO-TIP: If you need to clean the *original* data file, you should complete data cleaning tasks before loading the data into Tableau.

Hiding Columns

Most data analysis projects require some amount of data cleaning. Sometimes, you have columns in your dataset that you do not need for your project. In Tableau, you can hide them by clicking the drop-down arrow (or right-clicking the column header area) and selecting *Hide*.
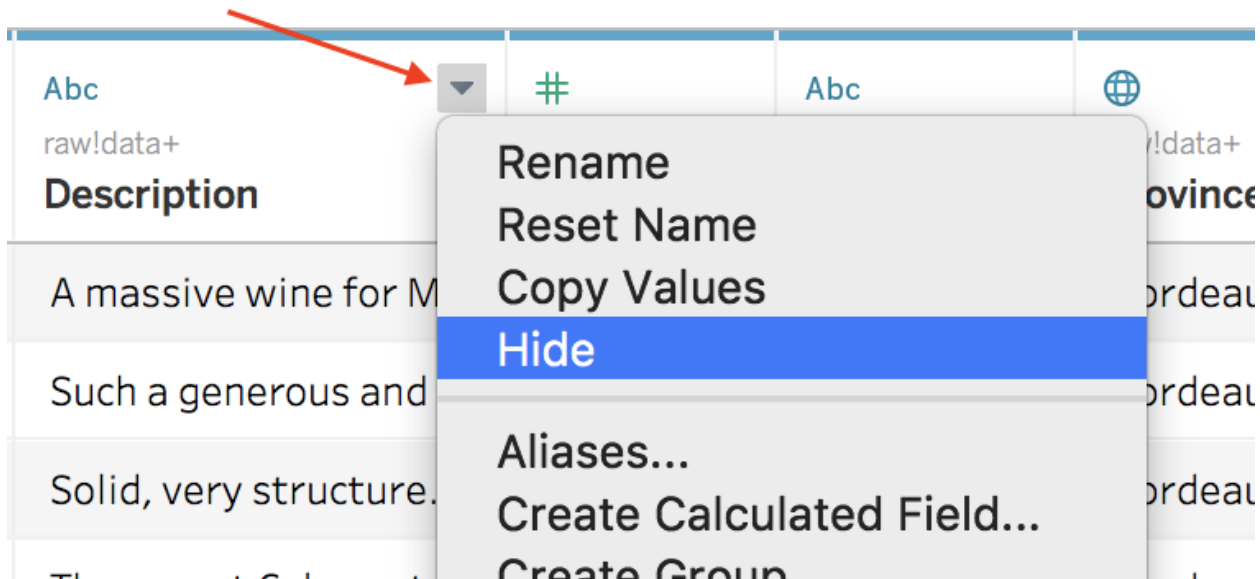
Drop-down arrow found in the header area leads to menu with the Hide column option. As mentioned earlier, you are not actually deleting the columns, but you are instead filtering them out from the workbook file.

If you need to "unhide" the columns later down the road, then all you need to do is return to the Data Source page, and click on the checkbox for *Show hidden fields,* as shown below. You should see the hidden columns grayed out, but visible. Then, you can click on the drop-down arrow for the column and select *Unhide*.
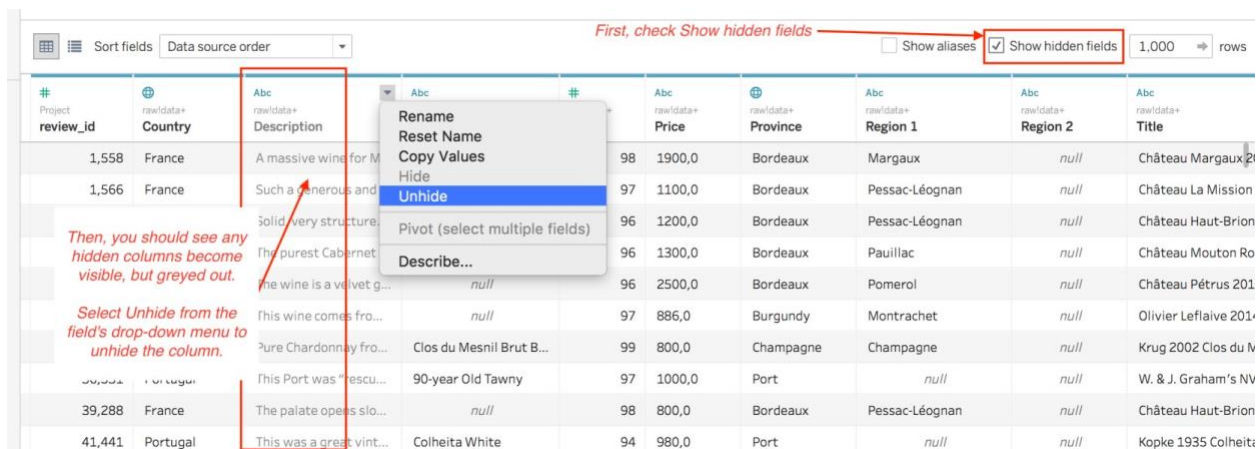
Click Show hidden fields checkbox to see and unhide and hidden fields.

After combining data, Tableau creates columns in the dataset. When you use a union to combine data from different tables, Tableau creates **two columns** (Sheet and Table Name) to inform you what the original data source is for the row of data. When you use a JOIN, Tableau brings in the **common column from all the data tables** involved. In our case, we have the *T code* column from the left table and the *T code column* from the right (*wine_experts*) table.

We needed both T code columns to carry out the join (remember the yellow join column from the previous chapter?). Now that we have joined the tables using the T codes, however, we no longer need these fields for our analysis. Let's go ahead and hide them as well as others that are not useful.

Can you identify those columns? Answers below!
Hide the columns listed below:

- *Sheet*
- *Table Name*
- *T code*
- *T code (wine!experts)*
- *Description*
- *Designation*
- *Region 2*

You should have only the columns displayed below in Section E of the Data Source page:

#	⊕	#	Abc	⊕	Abc	▼	Abc	Abc	Abc	Abc	Abc
Project	raw!data+	raw!data+	raw!data+	raw!data+	raw!data+	⊏̄	raw!data+	raw!data+	raw!data+	wine!experts	wine!experts
review_id	Country	Points	Price	Province	Region 1		Title	Variety	Winery	Taster Name	Taster Twitter Handle
1,558	France	98	1900,0	Bordeaux	Margaux		Château Mar...	Bordeaux-st...	Château Mar...	Roger Voss	@vossroger
1,566	France	97	1100,0	Bordeaux	Pessac-Léognan		Château La M...	Bordeaux-st...	Château La ...	Roger Voss	@vossroger
1,571	France	96	1200,0	Bordeaux	Pessac-Léognan		Château Hau...	Bordeaux-st...	Château Hau...	Roger Voss	@vossroger

Remaining columns after hiding several fields in Section E.

Set Data Types

Tableau is pretty good at inferring the correct data types for your columns. However, sometimes they don't get it right, or you need to correct the data type that was set. For example, let's take a look at the *review_id* column. You can see that it has been set as a numerical data type by looking at the data type icon in the upper-left area of the header, as shown below.

#

Project

review_id

1,558

1,566

1,571

1,575

15,840

30,110   Data type icon for the review_id field.

If you take some time to understand the *review_id* field, you will see that it contains the unique row ID for the rows in the dataset. Even though the IDs are a series of numbers, they should be treated as identifiers for the rows of data and ***not*** as data values that can be aggregated. So, the *review_id* field should be set as a **string** field instead of a numerical data type. Let's fix that!

First, click on the data type icon for the *review_id* field, and you should see a menu of options.

- The ***numerical (decimal)*** data type is for fields that contain numerical values with decimal points. It is also known as the float data type.
- The ***numerical (whole)*** data type is for fields that contain whole number values. It is also known as the integer data type.
- The ***date and time*** data type is for fields that contain timestamps.

- The ***date*** data type is for fields that contain dates.
- The ***string*** data type is for fields that contain text (string characters).
- The ***Boolean*** data type is for fields that contain one of two possible values such as 0, 1, True or False.
- The ***geographic role*** data type is for geographical data.

Then, select the *String* option, and it is as easy as that! Notice how the values in the *review_id* column no longer have the commas displayed. This is because Tableau sees strings instead of numbers, which is correct for this column of codes (i.e., row IDs).

Now, look at the *Country* column. The data type icon displays a *globe,* which represents the *Geographic data type*. One of the powerful features of Tableau is the ability to visualize data onto a map using different types of geographic data.

Besides longitude and latitude coordinates, Tableau can infer the following geographic data values:

⊕ ▼	#	Abc	⊕	Abc
		raw!data+	raw!data+	raw!
		Price	**Province**	**Reg**

	Number (decimal)		1900,0	Bordeaux	Ma
	Number (whole)				
	Date & Time		1100,0	Bordeaux	Pes
	Date				
✓	String		1200,0	Bordeaux	Pes
	Boolean				
			1300,0	Bordeaux	Pau
✓	Default				
	Geographic Role ▶		None		or
	France	97	Airport		Mc
	France	99	Area Code (U.S.)		ha
			CBSA/MSA (U.S.)		
	Portugal	97	City		
	France	98	Congressional District (U.S.)		es
			✓ Country/Region		
	Portugal	94	County		
			NUTS Europe		
	France	97	State/Province		es
			ZIP Code/Postcode		
			Create from ▶		

The available options for geographic roles in Tableau.

As long as your dataset has a column containing any of those geographic data types, you will be able to create map visualizations with your data. In a later chapter, we will discuss the *Create from* option.

Replace String Values, Such as Commas

It looks like the rest of the columns have the correct data type, except the Price column. The data type for that column is set to a string instead of a numeric type. You can also see that there are commas in the data values. 🙃 Well, looks like we have some data cleaning to do!

The fact that Tableau has set the data type for the *Price* field to string ***and*** is displaying commas in the data values tells us that there are ***actual*** commas as string characters.
This means that we will have to address those commas, or Tableau will not be able to infer the numeric value correctly. In fact, if you try to set the data type for the *Price* field as Number (decimal) before addressing the commas, Tableau will convert the data values incorrectly.

What we need to do is:

1. Create a new column (also known as a calculated field) from Price, and remove the commas using a Tableau function called REPLACE().
2. Change the data type of the new column to Number (decimal).

Let's go ahead and rename the *Price* column to *Price_old* (we will eventually hide it). To rename a column, you can either double-click on the field name or select the *Rename* option from the drop-down menu for the field.

Now to create the new column, click open the drop-down menu for the *Price_old* column, and select the *Create Calculated Field...* option, as shown below.

Abc	▼	⊕	Abc

raw!data+

Price_old

	Rename
	Reset Name
1900,0	Copy Values
	Hide
1100,0	
	Aliases...
1200,0	**Create Calculated Field...**
	Create Group...
1300,0	Split
2500,0	Custom Split...
886,0	Pivot (select multiple fields)
800,0	Describe...

Select the Create Calculated Field... option from the drop-down menu for Price_old.

A window will pop up on your screen. There, you will use the first area to change the name of the new column from *Calculation1* to *Price*, as shown below. In the second area, type out the Tableau code or logic that is needed to create the new column. We will be using the Tableau function called `REPLACE` with the *Price_old* field to create the new column. Tableau has already added the notation for the *Price_old* column (which we are basing our new column on) in the second area.

When you are using Tableau code, column names are case-sensitive and need to be enclosed in square brackets: **[**Column Name example**]**

Recognized columns will appear in an orange font. This confirms that your code is properly referencing the columns in your dataset. If your column names are **not orange** in your
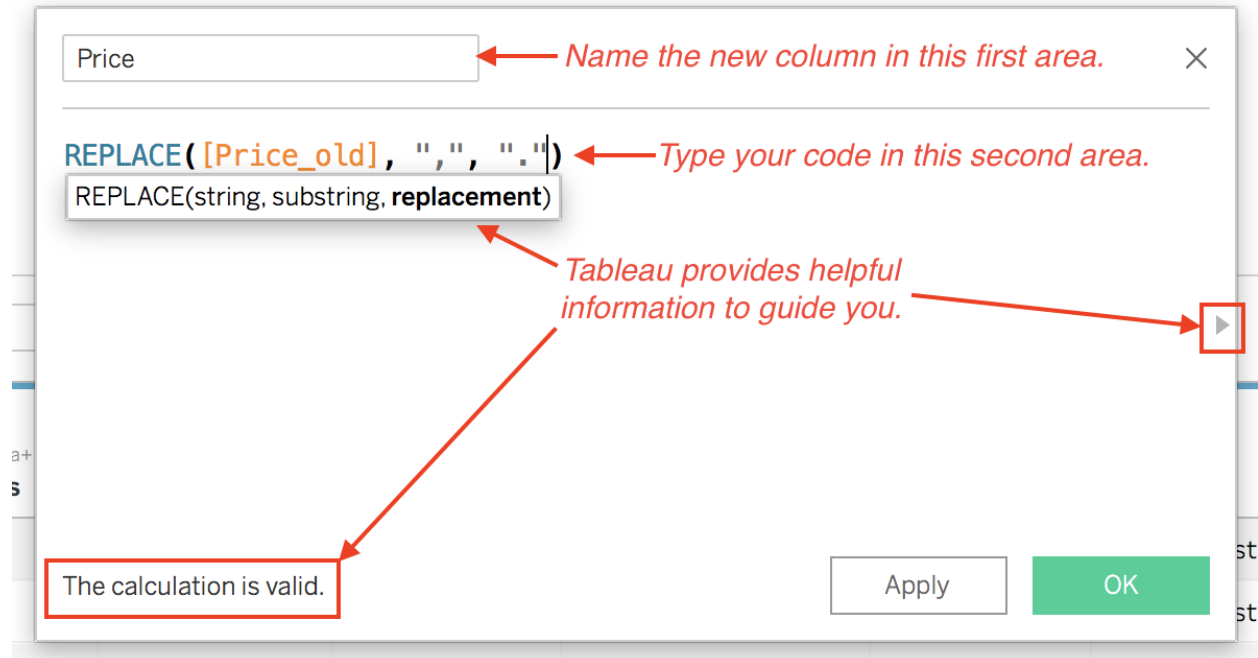
code, as shown below, then you will need to fix that reference (for example, you may have a 'typo' of some kind).

Put your cursor at the start of the line of code, and type out the following:

```
REPLACE([Price_old], ",", ".")
```

As you are typing, you can see that Tableau will provide some helpful information to guide you.

Dialogue window in which you write Tableau code to create a new column.

This `REPLACE()` function is telling Tableau to go to the data value in the `[Price_old]` column, find a comma substring character `","` and replace it with a decimal point `"."` (or rather, a period).

Make sure there are no errors in your code and that the calculation is valid (lower-left corner of the window). Then, click *OK* or *Apply*.

Now you should be able to set the *Price* column as a Number (decimal) data type, and Tableau will be able to convert the data values correctly. Take a look below at the difference between the *Price_old* column and the new *Price* column.

Abc	=#
raw!data+	Calculation
Price_old	**Price**
1900,0	1900.00
1100,0	1100.00
1200,0	1200.00
1300,0	1300.00
2500,0	2500.00
886,0	886.00
800,0	800.00
1000,0	1000.00
800,0	800.00

New price

column set as a Number (decimal) data type compared to the Price_old column.

Go ahead and hide the *Price_old* column since we do not need it anymore. At this point, I highly recommend that you save your work!

If you try to save your work while you are in the Data Source page, you might receive the following message from Tableau, which is preventing you from saving your work:
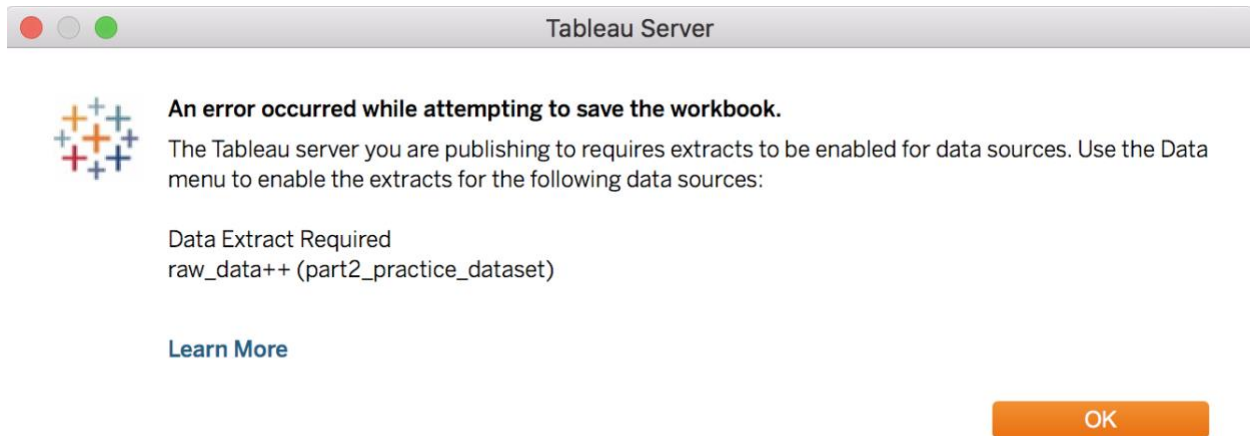
Save error that occurs when saving from the Data Source page.

To resolve this issue, click *OK* for that error message and navigate to Sheet 1, which we created in the previous chapter. From that screen, you will be able to save your work without an error message.

Remember Section F?

Let's end this chapter with a discussion about Section F of the Data Source page (lower-left area), shown below.
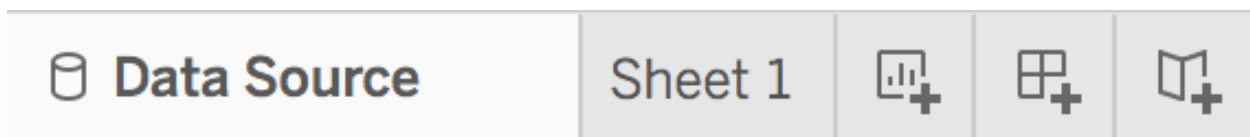
Section F of the Data Source page.

- The *Data Source* tab will take you to the Data Source page.
- Any sheets that you have created in your Tableau workbook will appear after the Data Source button. You should see a tab for Sheet 1, which will take you to that sheet.
- Then, you have three different **ADD icons**: *New Worksheet*, *New Dashboard*, and *New Story***.**

When you select the *New Worksheet* icon (the first ADD icon), Tableau will create a new blank worksheet. We will use the *New Dashboard* and *New Story* icons in Part 3 of this course. For now, we will only be using the *New Worksheet* icon.

Summary

- Most data analysis projects require some amount of data cleaning.
- Tableau does **not** change the **original data files** that are used to load data into the Tableau workbook.

- Tableau does, however, allow you to:
  - Hide (or filter out) columns.
  - Set data types.
  - Replace data types by creating calculated fields.